Biochem. J. (1995) 308, 89–96 (Printed in Great Britain)

Isolation and characterization of cDNA clones encoding pig gastric mucin

Bradley S. TURNER,* K. Ramakrishnan BHASKAR,*‡ Margarita HADZOPOULOU-CLADARAS,*† Robert D. SPECIAN† and J. Thomas LAMONT*

*Section of Gastroenterology, Evans Department of Clinical Research, Section of Molecular Genetics, Boston University Medical Center, 88 East Newton Street, Boston, MA 02118 and †Department of Cellular Biology, School of Medicine in Shreveport, Louisiana State University Medical Center, 1501 Kings Highway, Shreveport, LA 71130, U.S.A.

Polyclonal antibodies raised to deglycosylated pig gastric mucin were used to screen a cDNA library constructed with pig stomach mucosal mRNA. Immunocytochemistry indicated that the antibody recognizes intracellular and secreted mucin in surface mucous cells of pig gastric epithelium. A total of 70 clones producing proteins immunoreactive to this antibody were identified, two of which (PGM-2A,9B) were fully sequenced from both ends. Clone PGM-9B hybridized to a polydisperse mRNA (3–9 kb) from pig stomach, but not liver, intestine or spleen, nor to mRNA from human, mouse, rabbit or rat stomach. Sequence analysis indicated that PGM-9B encodes 33 tandem repeats of a 16-amino-acid consensus sequence rich in serine (46%) and threonine (17%). Using the restriction enzyme *Mwo*I, which has a single target site in the repeat, it was demonstrated that PGM-9B consists entirely of this tandem repeat. Southern-blot analysis indicated that the repeat region is contained in a 20 kb *Hin*dIII–*Eco*RI fragment, and *Bam*HI digestion suggested that most of the repeats are contained in a 10 kb fragment. *In situ* hybridization with an antisense probe to PGM-9B showed an intense signal in the entire gastric gland. Clone PGM-2A also contains the same repeat sequence as 9B, but, in addition, has a 64-amino-acid-long non-repeat region at its 5′ end. Interestingly the non-repeat region of PGM-2A has five cysteine residues, the arrangement of which is identical with that reported for human intestinal mucin gene *MUC2*.

INTRODUCTION

Mucus provides a lubricating and defensive barrier between the epithelium and the external environment of several mammalian organs, including the lung, stomach, intestine and the uterus. Mucins, epithelial glycoproteins, are the major secretory protein of mucus, and mucins from different organs share common structural features that contribute to the gel-forming and protective properties of the mucus layer [1–4]. A typical mucin has a very high molecular mass (> 2000 kDa) and consists of a linear peptide core (20%, by weight) with radially arranged oligosaccharide chains 3–20 sugar residues long (80% by weight), giving it a 'bottle brush' configuration [5].

The mammalian stomach is unique in that its epithelial lining is nearly constantly exposed to pH 2 or less as a consequence of HCl secretion. Mucin gel, present as a 100–400-μm-thick layer on the surface of the epithelium has been proposed as a protective diffusion barrier against luminal HCl. We have recently demonstrated [6] that the phenomenon of 'viscous fingering' offers a plausible physico-chemical mechanism by which the mucus gel layer can protect the underlying gastric epithelium while allowing the secretion of acid. Essential to this protective mechanism is the profound increase in the viscosity of gastric mucin produced by the secreted acid [7]. Our previous studies [7], as well those of others [8], have shown that the protein backbone (apomucin), although accounting for less than 25% of the glycoprotein mass, is essential for the formation of the mucin polymer that is responsible for the viscous properties of mucus. Lipid binding by gastric mucin [9], which contributes to the protective properties of gastric mucus [10], may also occur in the hydrophobic regions of apomucin.

The structure of gastric apomucin has eluded investigators because of its large size and extensive glycosylation. Recent advances in efficient chemical deglycosylation [11,12] and molecular cloning and sequencing have led to the characterization of several cDNA clones encoding parts of apomucin from various epithelial organs [13–17]. Our goal in the present study was to isolate a cDNA clone for pig gastric mucin in order to eventually obtain the complete amino acid sequence of apomucin. We employed polyclonal antibodies raised against deglycosylated pig gastric mucin (PGM) to screen a cDNA library constructed using mRNA isolated from pig stomach mucosa. A total of 70 positive clones were identified; in the present paper we describe the isolation and characterization of two of these cDNA clones, namely PGM-2A and -9B.

Part of this work was presented in abstract form at the Annual Meeting of the American Gastroenterological Association, held in New Orleans, LA, U.S.A., in May 1994 [17a].

EXPERIMENTAL

Mucin purification

PGM was isolated from pig stomach mucosal scrapings by size-exclusion chromatography in a Sepharose CL-4B (5 cm × 100 cm; Pharmacia, Piscataway, NJ, U.S.A.) column and purified by density-gradient ultracentrifugation in CsCl (42%, w/w) following standard procedures [7–9].

Deglycosylation of PGM

Deglycosylated PGM (PGM-HF) was prepared by deglycosylation of highly purified PGM using anhydrous HF (Peninsula

Abbreviations used: PGM, pig gastric mucin; PGM-HF, deglycosylated PGM; poly(A)$^+$, polyadenylated; pfu, plaque-forming unit(s); vWF, von Willebrand factor.

‡To whom correspondence should be sent.

The sequences described in this paper have been submitted to the GenBank Nucleotide Sequence Database and have been assigned the accession numbers U10281 for mucin clone PGM-9B and U12768 for PGM-2A.

Laboratories, Belmont, CA, U.S.A.) [11]. The amino acid composition of samples hydrolysed at 110 °C in 6 M HCl was determined by using a Beckman 7300 analyser [18]. Amino-sugar content was also determined on the analyser, and neutral sugars were determined by the phenol/H_2SO_4 reagent [19].

Antibody production

Polyclonal antibodies to PGM-HF were produced in rabbits using standard procedures [20]. IgG fraction was purified from the rabbit antiserum by rivanol/$(NH_4)_2SO_4$ precipitation.

Immunocytochemistry

Cellular localization of gastric apo-mucin was examined by immunocytochemistry using a double-antibody staining technique [21,22]. The sections were blocked with BSA for 1 h, rinsed with PBS, then stained for 4 h with the anti-PGM-HF antibody at 1:50 or 1:200 dilution, rinsed in PBS and stained for 2 h in a second antibody (goat anti-rabbit IgG conjugated to fluorescein isothiocyanate) at 1:50 dilution. The controls were either exposed to second antibody alone or to a non-specific immune serum in place of the first antibody.

RNA isolation

Total RNA was prepared from tissues by the method of Chomczynski and Sacchi [23]. Poly $(A)^+$ mRNA was purified using oligo(dT)–cellulose (Stratagene [24].)

cDNA library construction

First-strand synthesis was accomplished by denaturing 5 μg of poly$(A)^+$ mRNA in 10 mM methylmercuric hydroxide, 70 mM 2-mercaptoethanol [25], priming with 0.1 μg of random hexamers, 1.3 μg of random nonamers and 1 μg of oligo-$(dT)_{12-18}$ using RNase H/Moloney-murine-leukaemia-virus reverse transcriptase (Life Technologies, Bethesda, MD, U.S.A.). RNA was degraded with *Escherichia coli* RNase H, and second-strand synthesis [26] accomplished by *E. coli* DNA polymerase I, blunting with T4 DNA polymerase. Hemiphosphorylated *Eco*RI–*Not*I–*Sal*I adapters were ligated using T4 DNA ligase and phosphorylated with T4 kinase. The resultant cDNA was size-selected using Sephacryl S-500 HR chromatography and fragment sizes were estimated by denaturing alkaline agarose electrophoresis in 1% agarose [27]. Fractions containing high-molecular-mass (> 0.5 kb) cDNA were precipitated with ethanol, dried, and ligated to *Eco*RI-digested, calf-intestinal-alkaline-phosphatase-treated lambda ZAPII (Stratagene) arms, and packaged into phage particles using GigapakII (Stratagene). The library was amplified once in SURE *E. coli* (Stratagene) on 150 mm × 15 mm NZY agarose plates at a density of 5×10^4 plaque-forming units (pfu)/plate, eluted into SM phage buffer (0.1 M NaCl/10 mM $MgSO_4$/50 mM Tris/HCl/0.1% gelatin, pH 7.5), made 7% with dimethyl sulphoxide and stored at −70 °C.

Library screening

The library was plated on SURE *E. coli* at a density of 2.5 × 104 pfu/150 mm × 15 mm NZY agar plate. Nitrocellulose membranes (137 mm; BA85; Schleicher and Schuell, Keene, NH, U.S.A.) saturated with isopropyl D-thiogalactopyranoside were used to obtain plaque lifts containing expressed proteins. These were screened using anti-PGM-HF antibody at a dilution of 1:2000 in 1% skim milk (Blotto [28]). Plaque-purified antibody-reactive clones were transformed into plasmid Bluescript SK(−) by *in vivo* excision using ExAssist helper phage and SOLR *E.coli* (Stratagene).

Northern- and slot-blot analysis

Purified RNA was subjected to denaturing electrophoresis [29] in 1% agarose, transferred to positively charged nylon membranes (Hybond-N^+; Amersham, Arlington Heights, IL, U.S.A.) [30] in 50 mM NaOH and covalently immobilized on the membrane by exposure to UV light. For slot-blot analysis, RNA samples were spotted on to positively charged nylon membranes and fixed as described above. RNA on the blots were prehybridized in 5 × SSC (1 × SSC is 0.15 M NaCl/0.015 M sodium citrate)/5 × Denhardt's, 50% formamide 0.1% SDS, 0.1 mg/ml salmon sperm DNA at 42 °C for 24 h, and hybridized in pre-hybridization solution containing 10% (w/v) dextran sulphate and cDNA probes at 1×10^6 c.p.m./ml, labelled with [^{32}P]dATP by random primer extension [31] and purified through Chroma-Spin 100 columns (Clontech, Palo Alto, CA, U.S.A.). Hybridized blots were washed twice at room temperature with 1 × SSC/0.1% SDS for 15 min, twice at room temperature with 0.25 × SSC/0.1%SDS for 15 min, once at 60 °C with 0.25 × SSC/0.1%SDS for 15 min, and examined by autoradiographic exposure of X-ray film (XOmat-AR5, Kodak) using intensifying screen (Cronex, du Pont, Wilmington, DE, U.S.A.).

Sequencing of cDNA clones

The sequencing strategy utilized nested deletion method [32] (Erase-a-base; Promega, Madison, WI, U.S.A.) of both strands of the insert. Supercoiled plasmid Bluescript SK(−) was purified by acid phenol extraction [33], digested with *Bst*XI and *Bam*HI for T3 deletions and with *Kpn*I and *Bsp*106 for T7 deletions, digested with exonuclease III at 30 °C at 45 s intervals yielding approx. 150 bp deletions. DNA was treated with S1 nuclease, gel purified in 1% SeaPlaque (FMC, Rockland, ME, U.S.A.) low-melt agarose, in 40 mM Tris/5mM sodium acetate, pH 7.6. Deleted bands were excised and ligated with T4 ligase directly in the melted gel [34], and used to transform competent [35] XL1-MRF *E. coli*. The deleted cDNA inserts in plasmid Bluescript SK(−) were purified by the Magic Miniprep alkaline lysis method (Promega), denatured with alkali [36], and sequenced by the Sanger dideoxy chain-termination method [37] employing Sequenase V 2.0 modified T7 DNA polymerase (United States Biochemical, Cleveland, OH, U.S.A.), T7 and T3 primers, [^{35}S]dATP, and 0.4–1.2 mm-wedge 6%-PAGE in 89 mM Tris/28.5 mM taurine/0.5 mM EDTA. The resulting sequences were aligned, translated into the corresponding amino acid sequences and analysed for restriction sites and repeat regions using Intelligenetics PC/GENE software (Intelligenetics, Mountain View, CA, U.S.A.). The sequences were comparedwith other known sequences in the Genbank and EMBL databases using the BLAST [38] e-mail file server of the National Center for Biotechnology Information at the National Library of Medicine (National Institutes of Health, Bethesda, MD, U.S.A.)

Partial digestions

PGM-2A and -9B inserts were obtained by complete digestion of the pBluscript vector with *Eco*RI and purification from 1 × TAE (40 mM Tris/20 mM sodium acetate/5 mM EDTA, pH 8.0)/1% low-melt agarose (SeaPlaque GTG), using Qiaex gel extraction method (Quiagen Inc., Chatsworth, CA, U.S.A.). The purified inserts were digested with *Mwo*I (0.5 unit/μg of DNA) at 37 °C, 10 μl aliquots removed from the digestion reactions at 5, 10, 20,

Table 1 Amino acid and carbohydrate composition of PGM and PGM-HF

Values for amino acids are expressed as residues/1000 residues and for sugars as mg of sugar/mg of protein. Abbreviations: nd, not determined because of interference with amino sugars; AA, amino acid.

Amino acid or sugar	Composition	
	PGM	PGM-HF
Asp	57	55
Thr	163	145
Ser	230	182
Glu	82	75
Pro	135	123
Gly	76	67
Ala	67	5
Val	nd	64
Cys	nd	8
Met	28	5
Ile	17	33
Leu	31	42
Tyr	17	16
Phe	nd	nd
His	18	16
Lys	22	21
Arg	37	30
Hex/AA	1.41	0.02
GlcNAc/AA	1.16	0.02
GalNAc/AA	0.62	0.03

Figure 1 Western-blot analysis of PGM and deglycosylated (PGM-HF) with anti-PGM-HF antibody

The antibody shows strong reactivity with PGM-HF (lane 1), extending over a broad range of molecular size (30–250 kDa). It also reacts with PGM (lane 2), but the reactivity is confined to the stacking gel, indicating that the antigen is an integral part of mucin.

40 and 80 min, resolved by PAGE on 4–20% pre-cast gradient gels (Bio-Rad Laboratories, Richmond, CA, U.S.A.) in 1 × TBE (89 mM Tris/89 mM boric acid/2 mM EDTA, pH 8.3) and bands revealed by staining with ethidium bromide.

Southern-blot analysis

Pig genomic DNA isolated from peripheral blood (Novagen, Madison, WI, U.S.A.) was subjected to digestion with the restriction endonucleases *Bam*HI, *Eco*RI, *Hin*dIII, *Kpn*I, *Pst*I, and *Sac*I. The enzyme digests were electrophoresed in 0.8% agarose, transferred to positively charged nylon membrane in 0.4 M NaOH [30], immobilized by exposure to UV light and hybridized to a ^{32}P-labelled probe derived from clone PGM-9B as described for Northern blots.

Figure 2 Immunohistochemistry of pig gastric mucosa using anti-PGM-HF antibody

Reactivity to anti-PGM-HF antibody was observed in surface mucous cells and also in the surface mucus gel, but not in neck and gland cells. Controls using secondary antibody alone or a non-specific immune serum in place of the first antibody did not show any staining. Magnification 120 ×.

In situ hybridization

A 3-month-old pig was perfused with 4% paraformaldehyde in PBS, and tissue specimens from the gastric fundus were collected and processed by the method described by Simmons et al. [39]. *In situ* hybridization was carried out using sense and antisense ^{35}S-labelled RNA probes transcribed form linearized DNA templates containing the full-length cDNA insert of clone PGM-9B.

RESULTS

Our aim in the present study was to isolate cDNAs specific to pig gastric apomucin by using antibodies raised to deglycosylated

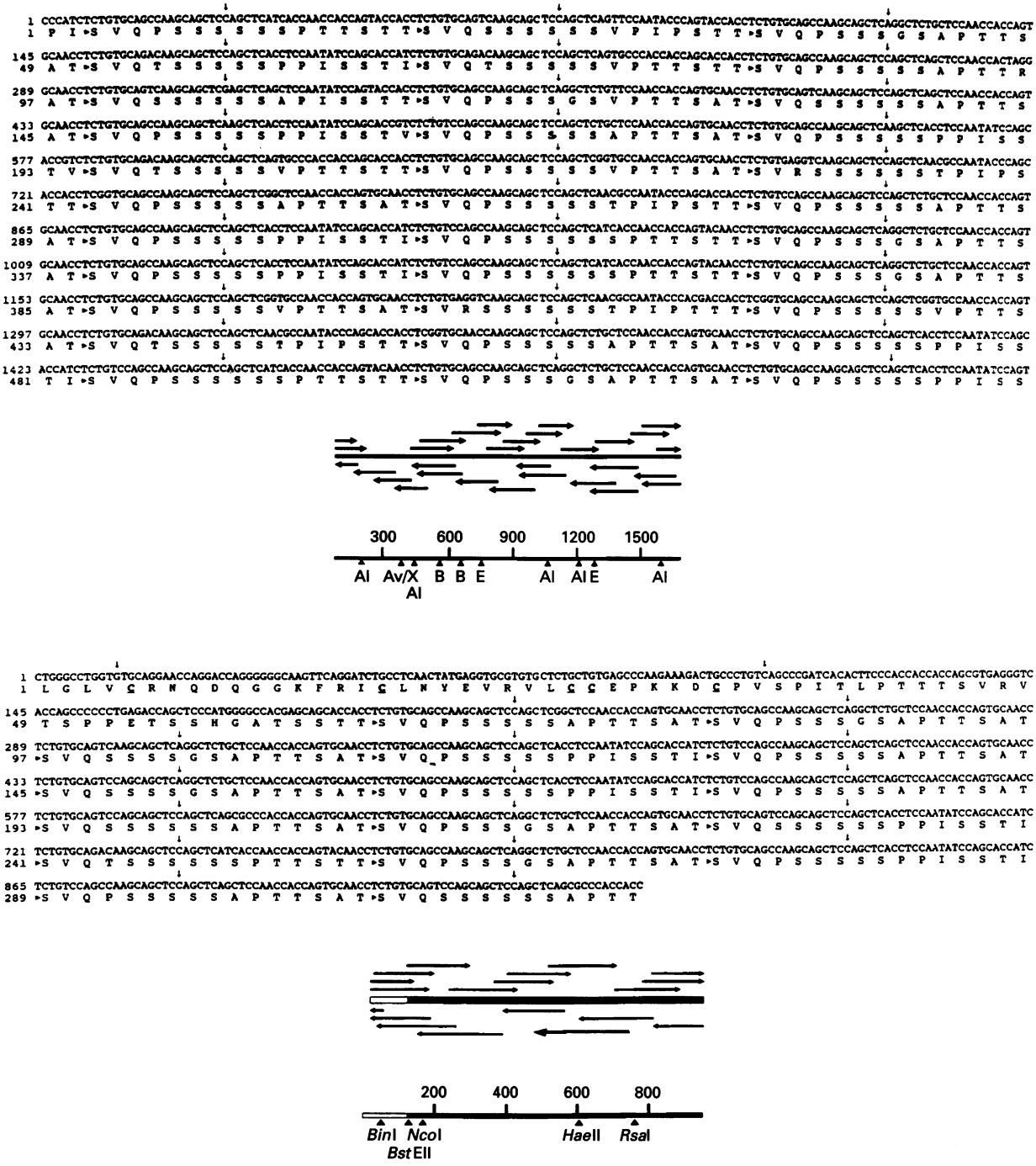

Figure 3 Sequence analysis of cDNA clones PGM-9B (top) and 2A (bottom)

Sequencing strategies, with the length and direction of individual sequencing reactions represented by arrows, are shown below the respective sequences. Key to restriction endonucleases: AI, *Alw*NI; Av, *Ava*I; B, *Bsm*AI; E, *Eco*NI; X, *Xho*I.

PGM to screen a pig gastric cDNA library . Examination of our mucin preparation by SDS/PAGE showed no significant low-molecular-mass contaminants: both silver and periodate/Schiff staining were confined to the stacking gel (result not shown). Further evidence of purity came from amino acid analysis, which showed the characteristic high preponderance of serine and threonine residues (Table 1).

Amino acid and carbohydrate compositions of native (PGM) and deglycosylated mucin (PGM-HF) are shown in Table 1. Over 95% of the sugars were removed by HF treatment, but the amino acid composition was essentially unchanged, except for a slight loss of serine (Table 1). On silver staining of SDS/PAGE gels, PGM-HF was seen as a broad smear in the separating gel (range 30–250 kDa).

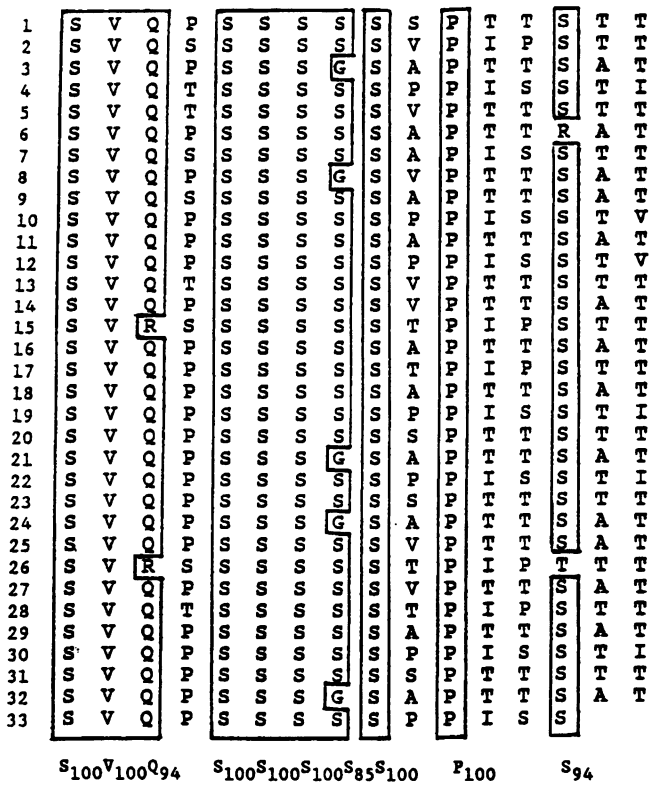

Figure 4 Aligned tandem repeats of cDNA clone PGM-9B

Boxed-in positions have > 85% conservation of amino acids.

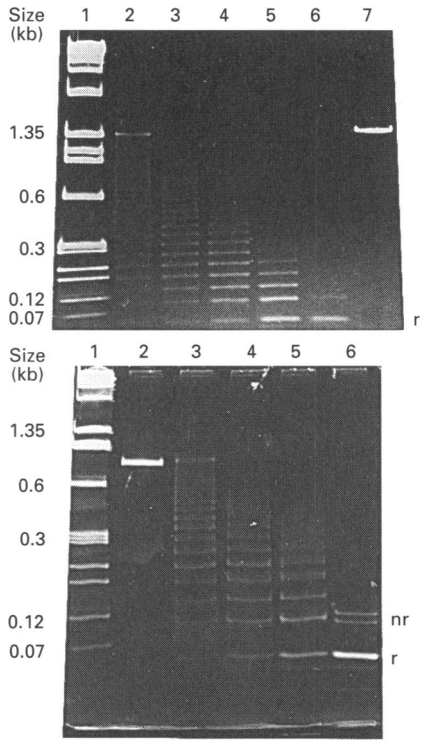

Figure 5 Partial digest with restriction enzyme *MwoI*

Top, PGM-9B: lane 1, size markers; lanes 2–5, sequential aliquots of digests; lane 6, after complete digestion; and lane 7, undigested 9B (1.58 kb). The single band of approx. 48 bp (lane 6) represents the repeat unit (r); the series of bands (lanes 2–5) arise from multiples of this sequence. Bottom, PGM-2A: PGM-2A (lane 2) also gives a band of 48 bp (r) after complete digestion (lane 6), but, in addition, two bands of approx. 90 and 95 bp are seen (nr); these correspond to the two fragments of the non-repeat region cleaved by the enzyme. The ladder of bands seen in intermediate digests (lanes 3–5) represent multiples of the repeat unit alone or in combination with fragments of the non-repeat region.

Apomucin antibodies

Anti-PGM-HF reacted strongly with native PGM also, but this reactivity was lost on Pronase treatment, implying that a reactive site is located in a proteinase-susceptible region. Treatment with dithiothreitol also abolished the reactivity of native PGM, further supporting the notion that an antigenic site is located in a peptide region. Western-blot analysis (Figure 1) indicated that the reactivity of PGM-HF to anti-PGM-HF extended as a smear through the separating gel (range 30–250 kDa), whereas the reactivity of native mucin to anti-PGM HF antibody was confined to the stacking gel, indicating that the antigenic site is an integral part of the mucin macromolecule. Anti-PGM-HF did not react with other proteins known to occur in the stomach, such as lysozyme, pepsin, intrinsic factor and albumin (results not shown).

Treatment of pig gastric mucosa with anti-PGM/HF resulted in intense staining of surface mucous cells and also of the surface mucus gel (Figure 2). No staining was observed in controls using secondary antibody alone or when a non-specific immune serum was substituted for anti-PGM-HF antibody. However, neck and gland mucous cells, which also secrete mucin, did not stain, indicating that the antibody reactivity is confined to the granular contents of surface mucous cells of pig gastric epithelium. The antibody stained mucin in some surface and gland mucous cells in pig trachea and also stained a very distinct subpopulation of goblet cells in the jejunum of pig, whereas most of the goblet cells were unstained. The antibody did not stain goblet cells in the pig ileum or colon, nor was there any cross-reactivity with liver (results not shown). The antibody did not cross-react with rat gastric goblet cells, suggesting that the antibody is also species-specific.

Isolation and characterization of cDNA clones

Anti-apomucin antibody was used to screen approx. 0.5×10^6 clones of the pig stomach cDNA library. A total of 70 plaques producing proteins immunoreactive to anti-PGM-HF were identified. Of these, ten were isolated as pure clones and have been partially characterized. They ranged in size from 0.6 to 2.5 kb. Two of these, clones PGM-2A and PGM-9B, have been sequenced in full and are described below. Partial sequence analysis indicated that the other eight purified clones were similar to either 9B or 2A.

The complete sequence of both strands of clone PGM-9B determined using an Exonuclease III nested deletion strategy (Figure 3, top) indicated that PGM-9B consists entirely of tandem repeats encoding a novel 16-amino-acid consensus sequence (SVQPSSSSSAPTTSTT) rich in serine (46%) and threonine (17%), which represent potential O-glycosylation sites. The tandem repeat sequences (Figure 4) show > 85% conservation of amino acids in all but six positions (4,10,12,13,15 and 16). It is noteworthy that serine, which is the predominant amino acid in pig gastric mucin (see Table 1), accounts for 46% of the

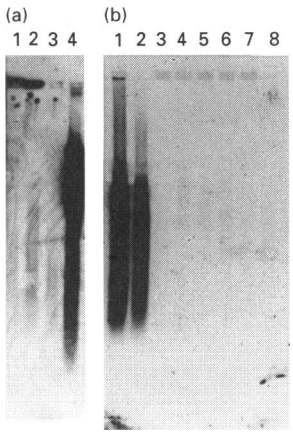

Figure 6 Northern-blot analysis of clone PGM-9B

(**a**) Lane 1, pig intestine; lane 2, liver; 3, spleen; and 4, stomach. Clone PGM-9B hybridized strongly to a large polydisperse (approx. 3–9 kb) mRNA from pig stomach, but not from pig liver, intestine or spleen. PGM-2A gave similar hybridization pattern (not shown). (**b**) Lane 1, pig stomach; lane 2, pig stomach; lane 3, pig liver; lane 4, mouse stomach; lane 5, rat stomach; lane 6, rabbit stomach; lane 7, human stomach; lane 8, gastric cell line AGS. PGM-9B hybridized strongly to RNA from pig stomach, but showed no hybridization to RNA from stomachs of other species tested or gastric cell lines.

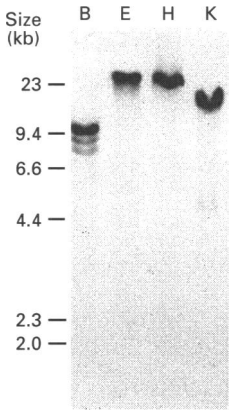

Figure 7 Southern-blot analysis of cDNA clone PGM-9B

Hybridization of probe derived from PGM-9B gave single bands of approximately the same size (> 20 kb) with *Eco*RI and *Hin*dIII digests, a single band of slightly smaller size with *Kpn*I and a triplet with *Bam*HI digest.

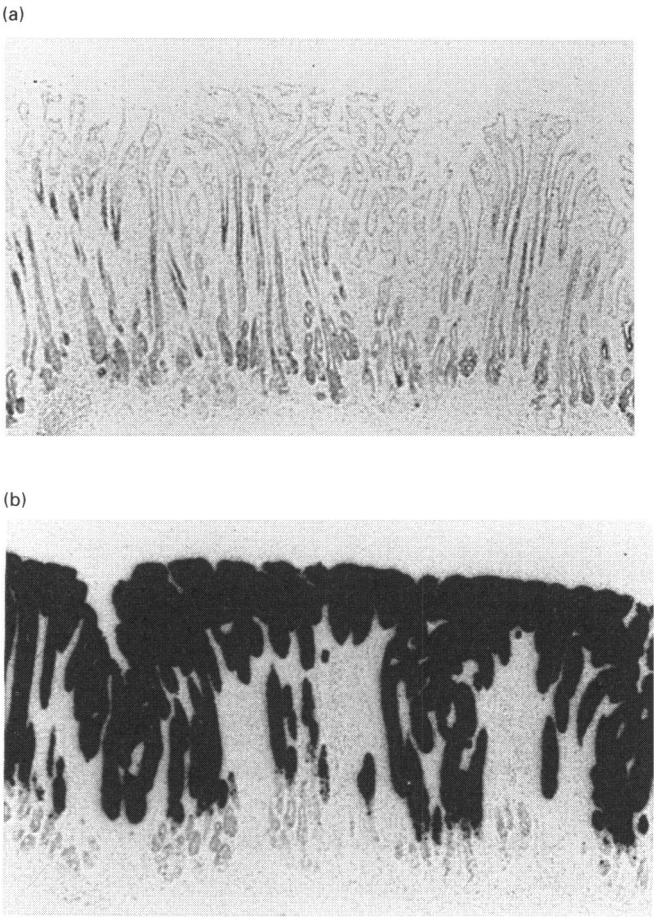

Figure 8 *In situ* hybridization using cDNA clone PGM-9B

Sense and antisense ^{35}S-labelled RNA probes transcribed from linearized DNA templates containing the cDNA insert of clone PGM-9B were hybridized to fresh pig stomach mucosa as described in the text. The antisense probe (**b**) gave a strong signal in the entire gastric gland, whereas no signal was observed with the sense probe (**a**). Magnification 45 ×.

average amino acid content of the tandem repeat. A search among known sequences in the Genbank and EMBL Nucleotide Sequence Databases [38] indicated that the sequence of clone PGM-9B is new.

Sequence analysis of clone PGM-2A (Figure 3, bottom) showed that it consists of 16 tandem repeats of the same 16-amino-acid consensus sequence as PGM-9B, but showed no overlap with 9B, indicating that it is from another part of the repeat region. In addition PGM-2A has a non-repeating sequence of 64 amino acids at its 5′ end. Interestingly the short stretch of non-repeat sequence has five cysteine residues, two of them in consecutive positions.

Examination of the repeat sequence of the two clones indicated the presence of a single target site (GC-N_7-GC) for the restriction enzyme *Mwo*I. Since PGM-9B consists entirely of repeats, complete digestion with the enzyme should yield the tandem repeat unit, whereas a partial digest will contain fragments representing multiples of the tandem repeat unit. As shown in Figure 5(a), complete digestion with the enzyme (lane 6) did in fact give rise to a single band of approx. 48 bp, which corresponds to the size of the tandem repeat unit. Aliquots of the digest taken at intermediate time points (lanes 2–5) showed a ladder of evenly spaced bands representing fragments containing multiples of the tandem repeat unit.

Clone PGM-2A has an additional *Mwo*I restriction site in the middle of the non-repeat region of PGM-2A, and therefore complete digestion with this enzyme would yield two fragments 90 and 95 bp long in addition to the 48 bp repeat unit. PAGE of *Mwo*I digests of 2A (Figure 5b) confirm that such indeed is the case.

The tissue distribution of the cDNA clones was examined by Northern-blot analysis of mRNA isolated from several different tissues. Both clones, PGM-2A and-9B, hybridized strongly to a large (approx. 3–9 kb) polydisperse pig stomach mRNA, but not to RNA from pig liver, intestine or spleen (Figure 6a), indicating that expression of both clones is tissue-specific.

The hybridization of PGM-2A is likely to be influenced mainly by the repeat structure, which is predominant, rather than the

short non-repeat sequence, thus resulting in a hybridization pattern similar to that of PGM-9B. In the following studies PGM-9B was therefore used as the representative clone.

PGM-9B showed no hybridization to mRNA from mouse, rat, rabbit or human stomach (Figure 6b), indicating that no sufficiently similar sequence is present in the other species tested. Slot-blot analysis indicated a progressive increase in intensity of the hybridization signal with an increasing amount of pig stomach mRNA. There was a strong signal with as little as 1 μg of pig stomach RNA, but no signal with as much as 10 μg RNA from mouse, rat, rabbit or human stomach, nor RNA from two human gastric cell lines, KATO III and AGS, further attesting to the specificity of hybridization (results not shown).

Southern-blot analysis of *Eco*RI and *Hin*dIII digests of pig blood genomic DNA with clone PGM-9B (Figure 7) indicated that clone PGM-9B is contained in fragments larger than 20 kb. *The Bam*HI digest gave rise to an intense band at approx. 10 kb and two weaker bands of slightly smaller size. The strong intensity of the 10 kb band suggests that most of the repeats are contained within this fragment.

In situ hybridization with ^{35}S-labelled antisense probe to a full-length insert of clone PGM-9B resulted in a very strong signal in the entire gastric gland of pig stomach epithelium, except for the base (Figure 8b), suggesting abundant presence of mRNA complementary to cDNA clone PGM-9B. In contrast, no staining occurred with the sense probe (Figure 8a), indicating specificity of the hybridization to clone PGM-9B.

DISCUSSION

We report here a novel cDNA clone (PGM-9B) which encodes a tandem repeat region of pig gastric mucin. Clone PGM-9B has all the features characteristic of mucin genes: (1) *in situ* hybridization indicates that the mRNA is localized in abundance in mucin secreting cells of the gastric surface epithelium; (2) it has a tandem repeat structure rich in serine and threonine, which are potential O-glycosylation sites; and (3) it hybridizes to a large, polydisperse mRNA in Northern-blot analysis. Since the antibody used to screen the library was raised to deglycosylated mucin and was shown to stain pig gastric mucous cells almost exclusively, we conclude that clone PGM-9B must encode a portion of pig gastric apomucin.

Northern- and slot-blot analysis indicated that clone PGM-9B is unique to pig stomach epithelium, since no hybridization to mRNA from various gastrointestinal and other tissues of different species was detected. The specificity of PGM-9B does not rule out the presence of other mucin gene homologues in the pig stomach, given the fact that a number of different genes appear to code for mucins, even in the same species, e.g. *MUC1–MUC6* [16,17].

The specificity of PGM-9B could be due to the fact that it encodes a tandem repeat region of pig gastric apomucin: studies of other mucin genes published so far seem to suggest that whereas non-repeat regions are conserved, tandem repeat sequences vary between tissues and species [17]. Tandem repeat sequences ranging in size from six [14] to 169 amino acids/unit [16] have been reported. There are also differences in the relative proportions of serine and threonine, which are potential glycosylation sites. The tandem repeat of another mucin gene cloned from the same species, namely pig submaxillary [40] bears no resemblance in size or sequence to PGM-9B; it is 81 amino acids long and its composition also differs from that of PGM-9B.

Interestingly, tandem repeat sequences of the same size as PGM-9B, namely 16 amino acids long, have been reported for a few other mucins. Porchet et al. have reported a 16-amino-acid tandem repeat sequence for the human tracheobronchial mucin MUC4 [41]. A 16-amino-acid consensus repeat sequence rich in serine and threonine has recently been identified in the mouse gastric mucin gene [42]. The human intestinal mucin gene, *MUC2*, also has a region coding for 16-amino-acid repeats within a larger, 39–40-amino-acid repeat [43]. Although the tandem repeats of MUC-2 and PGM-9B are different from one another, there appears to be some similarity when one takes into account conservative amino acid substitutions, such as between serine and threonine (see below).

Tandem repeats

```
PGM      S V Q P S S S S S A P T T S T T
         *     | * * * * *   |   | * |
MUC-2    T P P P T T T T T P P P T T T P
```

|, Identical amino acids; *, conservative substitutions

It remains to be established if such substitutions have any structural or functional significance.

Pig gastric mucin has a higher content of serine than threonine (the present results and [44]), whereas in other mucins, such as those from lung [45] and intestine [14], there is a higher threonine than serine content. Interestingly, the amino acid composition of the tandem repeat sequence of clone PGM-9B has a considerably higher content of serine (46%) than threonine (17%). Since PGM-9B encodes the serine- and threonine-enriched repeat region, the overall composition will be influenced by the relative proportion of these two amino acids in the repeat sequence. The higher content of serine than threonine in the tandem repeat sequence provides yet further support for the notion that PGM-9B encodes pig gastric apomucin.

The repeat sequence of clone PGM-9B (48 bp) is considerably smaller than the tandem repeat sequence of human gastric mucin reported recently [16], MUC6 (507 bp). The amino acid composition of the tandem repeat sequence of MUC6 is also strikingly different from that of PGM-9B. MUC6 has a higher content of threonine (30%) than serine (18%), whereas, in PGM-9B, the content of serine (46%) is considerably higher than that of threonine (17%). The difference could arise from the presence of more than one mucin gene, which would also explain differences between PGM-9B and MUC6 in their tissue specificities: whereas PGM-9B shows exclusive specificity to pig stomach epithelium, MUC6 shows equal intensity of hybridization to gall bladder and terminal ileum (Figure 6c; [16]). Clearly further studies on the full sequences of various apomucins are needed to resolve these issues.

In situ hybridization studies demonstrated that clone PGM-9B is specific to the surface epithelium of pig stomach and does not react with the underlying tissue. The fact that *in situ* hybridization extended over the entire gastric gland, unlike antibody staining, which was limited to the surface cells, may reflect post-translational processing differences between surface and gland mucin.

The size of pig gastric mRNA to which the clone hybridizes is considerably larger (9 kb) than the clone itself (1.58 kb), suggesting that clone PGM-9B is a part of a much larger mucin cDNA. This is also evident from Southern-blot analysis, which indicated that clone PGM-9B is part of a genomic DNA larger than 20 kb. If the repeat sequences are uninterrupted by introns, as in the case of MUC2 [43], our results from *Bam*HI digest indicate that the tandem repeat region can be as long as 10 kb. Thus the tandem repeat region of pig gastric apomucin alone can be over 3000 amino acids long.

Our studies have identified a second pig gastric mucin cDNA clone which contains, in addition to predominantly repeat sequences, a short stretch of the non-repeat region. The fact that PGM-2A has the non-repeat region at the 5′ end of the repeat structure suggests that this represents the 3′ end of the N-terminal non-repeat region and the beginning of the tandem repeat region of pig gastric mucin. Cysteine residues in the non-repeat region are of considerable functional importance, since they are essential for polymerization of mucin. Interestingly, the arrangement of cysteine residues in pig gastric mucin (PGM-2A) is identical with that reported for human intestinal mucin (MUC2), as shown below:

```
PGM-2A   N Q D Q G G K F R I C L N Y E V R V L C C E P K K D C
         * |       |   |   * ▌   * | * * | |  ▌▌   |      ▌
MUC2     D Q F G N G P F G L C Y D Y K I R V N C C W P M D K C
```

▌, Identical cysteines; |, other identical amino acids; *, conservative substitutions.

Gum et al. [46] have recently reported overall sequence similarity of the N-terminus of MUC2 to prepro-von Willebrand factor (vWF), which contains four identical domains which function in oligomerization and packaging into specific storage granules. It is noteworthy that oligomerization of vWF occurs at low pH, as we have recently reported for pig gastric mucin [7]. The similarity of MUC2 to vWF, and our finding of considerable similarity of PGM-2A to MUC2, suggests that the cysteine-rich non-repeating domain of PGM is involved in oligomerization, and possibly gelation, critical functions of this molecule. Screening of our pig gastric cDNA library with the non-repeat region of PGM-2A as radiolabelled probe will enable us to isolate overlapping clones encoding more of this functionally important region of pig gastric mucin.

This work was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (grant DK-28195 to J. T. L.) and a Grant-In-Aid (no. 93012420 to M. H. C.) from the American Heart Association.

REFERENCES

1 Allen A. (1978) Br. Med. Bull. **34**, 28–33
2 Allen A. (1981) Physiology of the Gastrointestinal Tract, vol. 1 (Johnson, L. R., Christensen, J., Grossman, M. I., Jacobson, E. D. and Schultz, S. G., eds.), pp. 617–639, Raven Press, New York
3 Strous, G. J. and Dekker, J. (1992) Crit. Rev. Biochem. Mol. Biol. **27**, 57–92
4 Carlstedt, I., Sheehan, J. K., Corfield, A. P. and Gallagher, J. T. (1985) Essays Biochem. **20**, 40–76
5 Schachter, H. and Roseman, S. (19??) in The Biochemistry of Glycoproteins and Proteoglycans (Lennarz, W. J., ed.), pp. 85–160, Plenum Publishing Corp., New York
6 Bhaskar, K. R., Garik, P., Turner, B. S., Bradley, J. D., Bansil, R., Stanley, H. E. and LaMont, J. T. (1992) Nature (London) **360**, 458–461
7 Bhaskar, K. R., Gong, D., Bansil, R., Pajevic, S., Hamilton, J. A., Turner, B. S. and LaMont JT. (1991) Am. J. Physiol. **261**, G827–G832
8 Pearson, J. P., Allen, A. and Vanables, C. (1980) Gastroenterology **78**, 709–715
9 Gong, D., Turner, B., Bhaskar, K. R. and LaMont, J. T. (1990) Am. J. Physiol. **259**, G681–G686
10 Lichtenberger, L. M., Romero, J. J., Kao, Y. J. and Dial, E. J. (1990) Gastroenterology **99**, 311–326
11 Mort, A. J. and Lamport, D. T. A. (1977) Anal. .Biochem. **82**, 289–309
12 Edge, A. S. B., Faltynek, C. R., Hof, L., Reichert, L. E., Jr. and Weber, P. (1981) Anal. Biochem. **118**, 131–137
13 Gum, J. R., Byrd, J. C., Hicks, J. W., Toribara, N. W., Lamoport, D. T. A. and Kim, Y. S. (1989) J. Biol. Chem. **264**, 6480–6487
14 Gum, J. R., Hicks, J. W., Lagace, R. E. et al. (1991) J. Biol. Chem. **266**, 22733–22738
15 Xu, G., Huan, L., Khatri, I. A., Wang, D., Bennick, A., Fahim, R. E. F., Forstner, G. G. and Forstner, J. F. (1992) J. Biol. Chem. **267**, 5401–5407
16 Toribara, N. W., Robertson, A. M., Ho, S. B., Kuo, W.-L., Gum, E., Hicks, J. W., Gum, J. R., Byrd, J. C., Siddiki, B. and Kim, Y. S. (1993) J. Biol. Chem. **268**, 5879–5885
17 Gum, J. R. (1992) Am. J. Res. Cell. Mol. Biol. **7**, 557–564
17a Turner, B. S., Bhaskar, K. R., Hadzopoulos-Cladaras, M., Specian, R. D., LaMont, J. T. (1994) Gastroenterology **106**, A60 (abstr.)
18 Moore, S. and Stein, W. H. (1963) Methods Enzymol. **6**, 819–831
19 Kabat, A. (1972) Methods Enzymol **28B**, 263–264
20 Hurn, B. A. L. and Chantler, S. M. (1980) Methods Enzymol. **70**, 104–112
21 Oliver, M. G., Wiggins, S. S., Specian, R. D. (1990) Trans. Am. Microsc. Soc. **109**, 205–212
22 Oliver, M. G. and Specian, R. D. (1991) Anat. Res. **230**, 513–518
23 Chomczyynski, P. and Sacchi, N. (1987) Anal. Biochem.**162**, 156–159
24 Aviv, H., Leder, P. (1972) Proc. Natl. Acad. Sci. U.S.A. **69**, 1408–1412
25 Payvar, F. and Schimke, F. (1979) J. Biol. Chem. **254**, 7636–7642
26 Gubler, U., Hoffman, B. J. (1983) Gene. **25**, 263–269
27 McDonnell, M. W., Simon, M. N., Studier, F. W. (1977) J. Mol. Biol. **110**, 119–124
28 Johnson, D. A, Gautsch, J. W., Sportsman, J.R, Elder, J. H. (1984) Gene Anal.Tech. **1**, 3–8
29 Davis, L. G., Dibner, M. D. and Battey, J. F. (1986) Basic Methods in Molecular Biology, p. 143, Elsevier, New York
30 Chomczynski, P. (1992) Anal. Biochem. **201**, 134–139
31 Feinberg, A. P. and Vogelstein, B. (1984) Anal. Biochem. **132**, 6–13
32 Henikoff, S. (1984) Gene **28**, 351–359
33 Zasloff, M., Ginder, G. D., Felsenfeld, G. (1978) Nucleic Acids Res. **5**, 1139–1152
34 Steggles, A. (1989) Biotechniques **7**, 241–242
35 Inoue, H., Nojima, H. and Okayama, H. (1990) Gene **96**, 23–28
36 Hattori, M. and Sakaki,Y. (1986) Anal. Biochem. **152**, 323–328
37 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. **74**, 5463–5467
38 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) J. Mol. Biol. **215**, 403–410
39 Simmons, D. M., Arriza, J. L., Swanson, L. W. (1989) J. Histotechnol. **12**, 169–181
40 Timpte, C. S., Eckhardt, A. E., Abernethy, J. L. and Hill, R. L. (1988) J. Biol. Chem. **263**, 1081–1088
41 Porchet, N., Van Cong, N., Dufosse, J., Audie, J. P., Guyonnet-Duperat, V., Gross, M. S., Denis, C., Degand, P., Bernheim, A. and Aubert, J. P. (1991) Biochem. Biophys. Res. Commun. **175**, 414–422
42 Shekels, L. L., Lyfrogt, C. T., Kieliszewski, M., Ho, S. B. (1994) Gastroenterology **106**, A178 (abstr.)
43 Toribara, N. W., Gum, J. R., Culhane, P. J., Lagace, R. E., Hicks, J. W., Petersen, G. M., Kim, Y. S. (1991) J. Clin. Invest. **88**, 1005–1013
44 Hase, T., Suoth, K. and Takahashi, K. (1992) Biomed. Res.**13**, 149–154
45 Bhaskar, K .R., and Reid, L. (1981) J. Biol. Chem. **256**, 7583–7589
46 Gum, J. R., Hicks, J. W., Toribara, N. W., Siddiki, B. and Kim, Y. S. (1994) J. Biol. Chem. **269**, 2440–2446

Received 22 July 1994/13 December 1994; accepted 14 December 1994